Int J Legal Med (2001) 114:204–210

ORIGINAL ARTICLE

Peter Gill

An assessment of the utility of single nucleotide polymorphisms (SNPs) for forensic purposes

Received: 11 May 1999 / Accepted: 27 September 1999

Abstract This paper assesses the use of single nucleotide polymorphisms (SNPs) for forensic analysis. It demonstrates that relatively small arrays of approx. 50 loci are comparable to existing short tandem repeat (STR) multiplexes. A quantitative test, however, is a prerequisite for mixture interpretation. In addition, as the mixture proportion becomes low, it will be necessary to distinguish between the allele and background. Relatively small biallelic arrays are also suitable to distinguish between closely related individuals such as brothers.

Keywords SNP · Polymorphisms · Arrays · Multiplexes · Simulation · Biochips

Introduction

There is increasing interest in the use of biallelic markers or single nucleotide polymorphisms (SNPs) for forensic purposes (Syvanen et al. 1993). Several formats have been used for PCR-based biallelic assays: the reverse dot blot (Saiki et al. 1988) applied to HLA DQ-alpha and Polymarker systems, microtitre-based formats (Kostyu et al. 1993) and finally microfabricated arrays on glass (Southern et al. 1992, 1994; Guo et al. 1994). The latter are of special interest since the potential exists to build arrays consisting of hundreds of loci. This paper specifically explores the potential of biallelic arrays, particularly with respect to the analysis of mixtures. All of the platforms are non-electrophoretic.

A crucial aspect of forensic DNA typing is the interpretation of mixtures (Evett et al. 1991; Weir et al. 1997). Until recently, statistical interpretation of mixtures has proceeded without considering differences in signal strength of heterozygotes at a locus. Evett et al. (1998), Clayton et al. (1998) and Gill et al. (1998) reported methods to interpret mixed STR profiles based on identification of the allele peak areas. Although intended for STR (electrophoretic) analysis, the principles can be extended to encompass biallelic loci on non-electrophoretic media.

P. Gill
The Forensic Science Service, Trident Court,
2960 Solihull Parkway, Birmingham Business Park,
Birmingham B37 7YN, UK

How large does an array need to be?

Typically an array of biallelics could comprise several hundred loci that are typed from a single individual. In this paper I consider relatively small arrays of 50–150 loci. Excluding the possibility of genetic 'nulls', a consideration of each locus in turn must fall into one of two categories – either one allele will be visible or two alleles will be seen. The notation A, B is used to denote the two alleles where a and b denote their respective frequencies – only AA, AB or BB are possible (because $a + b = 1$, all formulae could be expressed solely in terms of a). If a microfabricated array consists of n loci, the match probability can be approximated by making the simplified assumption that the frequency of A (a) is constant for every locus in the array. For n different loci in an array, the number of AA genotypes is a^2n, the number of AB genotypes is $2abn$ and the number of BB genotypes is b^2n. For example if $n = 100$ and $a = 0.5$, then 50 loci will be heterozygote and 50 will be homozygote (AA and BB in equal proportion). Therefore, the match probability across the entire array is $(a^2)^{25} \times (2ab)^{50} \times (b^2)^{25}$. If estimated as a likelihood ratio (LR_n):

$$LR_n = \left(\frac{1}{a^2}\right)^{a^2n} \times \left(\frac{1}{2ab}\right)^{2abn} \times \left(\frac{1}{b^2}\right)^{b^2n}$$

Figure 1 shows simulations for arrays ranging from 50–150 loci. A relatively small array of 50 gives likelihood ratios equivalent to approximately 12 STRs over a wide range of $a > 0.2 < 0.8$. Note that the plots in Fig. 1 are symmetrical, so that the $LR_{(a = 0.8)}$ is the same as $LR_{(a = 0.2)}$.

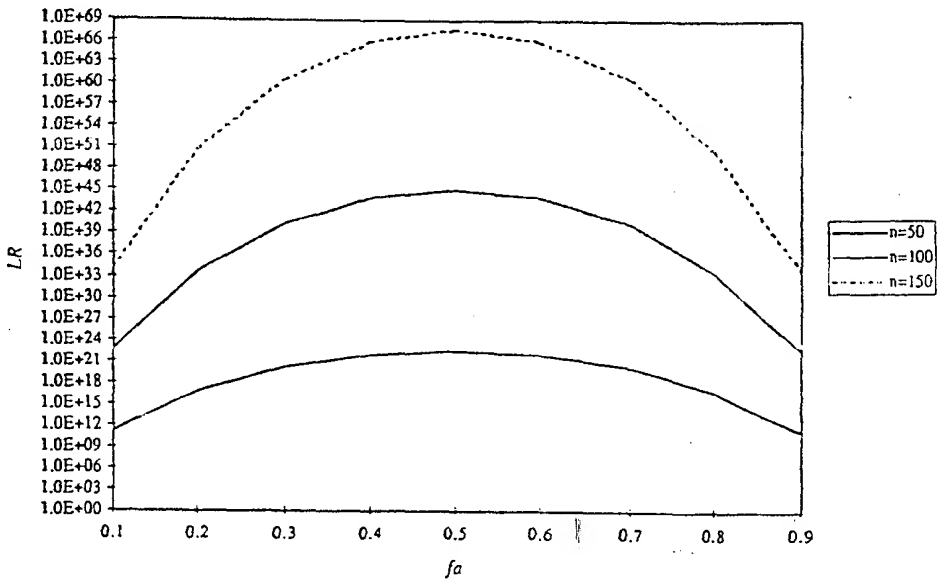

Fig. 1 Estimates of LR_n from arrays of n loci, assuming fa is constant across the set

Analysis of mixtures

Assuming two contributors to the mixture, if one allele shows then both must be homozygous for the same allele (*AA*,*AA* or *BB*,*BB*).

If there are two alleles visible, and assuming that there are two contributors to a mixture, (suspect S and victim V, respectively) then the following genotype combinations are possible: *AA*,*AB*; *AA*,*BB*; *AB*,*BB*; *AB*,*AB* and all of the reverse possibilities (Weir et al. 1997). This makes a total of nine possible genotype combinations (m = 1...9), all of which may be represented in a mixture. Given a normal outbreeding population, the proportion of observations of all of the above mixture types can be estimated given a.

Contributors to the mixture are the suspect and an unknown individual

For example, suppose that a blood stain is retrieved from a crime scene and the genotypes are consistent with a combination of the suspect (S) with an unknown individual (U).

We consider the following conditions in the likelihood ratio:

C: Contributors were the suspect and unknown

$\bar{C}$: Contributors were two unknown individuals

For each locus, calculation of the likelihood ratio depends upon the genotype of the suspect and the alleles observed in the mixture and there are three broad categories to consider.

Category 1

The suspect is homozygous *(AA)* and the mixture is *AB*

(U) must be either *AB* or *BB*

$$C = 2ab + b^2$$

$$\bar{C} = 6a^2b^2 + 4a^3b + 4ab^3$$

$$LR = (2ab + b^2)/(6a^2b^2 + 4a^3b + 4ab^3)$$

Category 2

The suspect is heterozygous *(AB)* and the profile is *AB*.

(U) must be *AA*, *AB* or *BB* and:

$$C = (a + b)^2$$

$\bar{C}$ is the same as in category 1 above.

Category 3

The suspect is homozygous *(AA)* and the profile shows just one allele. (U) is *AA* and the LR is $1/a^2$.

A complete list of numerators and denominators is given in Table 1. The proportion of an array of n loci having a particular mixture type *(m)* is *fm*: Each locus has $mp = 9$ possible mixture genotype combinations each (listed in Table 1).

The total $\overline{LR_n}$ of a mixture in an array of n loci is:

$$\overline{LR_n} = \prod_{m=1}^{mp} LR^{(fm \times n)}$$

Simulation of typical (average) mixture statistics on the combined $\overline{LR_n}$ for any number of biallelic loci was carried out under the simplified assumption that the allele proportion *(a)* for each locus is the same across loci (Fig. 2). $\overline{LR_n}$ maximises when a is high (0.8) or low (0.2). A battery of 50 loci with frequencies of alleles ranging between 0.1–0.9 will give a minimum LR of 10^4.

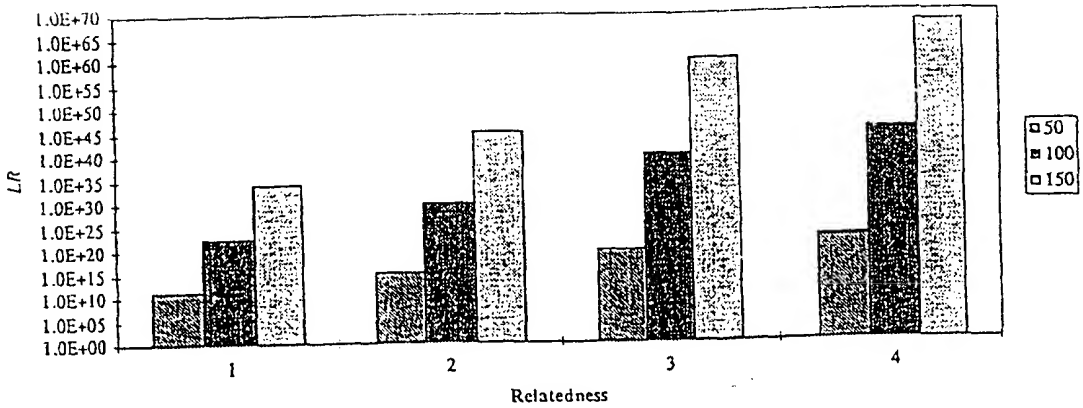

Fig. 7 A comparison of likelihood ratios from arrays of 50, 100 and 150 loci, respectively, under the assumption that the suspect and perpetrator are related. The allele proportion *(fa)* is 0.5 . On the *X-axis*: *1)* full brothers, *2)* father and son, *3)* first cousins, *4)* unrelated

tion can still proceed provided that cumulative probability functions can be used to estimate *p(null)*. Interpretation of more than two individuals contributing to a mixture will present a major challenge. Independence assumptions have not been assessed in this paper; however, it is inevitable that due consideration will be needed with large arrays.

Currently, the greatest problem in developing useful SNP arrays for forensic use is not related to statistical issues, rather, the problems are biochemical. Making a large balanced multiplex of ca. 50 loci from less than 1 ng of genomic template is indeed a daunting prospect.

References

Clayton TM, Whitaker JP, Sparkes R, Gill P (1998) Analysis and interpretation of mixed forensic stains using DNA STR profiling. Forensic Sci Int 91:55–70

Evett IW, Weir BS (1998) Interpreting DNA evidence. Sinauer Associates, Sunderland, Mass.

Evett IW, Buffery G, Willot G, Stoney DA (1991) A guide to interpreting single locus profiles of DNA mixtures in forensic cases. J Forensic Sci Soc 31:41–47

Evett IW, Gill P, Lambert JA (1998) Taking account of peak areas when interpreting mixed DNA profiles. J Forensic Sci 43: 62–69

Gill P, Sparkes R, Pinchin R, Clayton T, Whitaker J, Buckleton J (1998) Interpreting simple STR mixtures using allele peak areas. Forensic Sci Int 91:41–53

Guo Z, Guilfoyle RA, Thiel AJ, Wang R, Smith LM (1994) Direct fluorescence analysis of genetic polymorphisms by hybridization with oligonucleotide arrays on glass supports. Nucleic Acids Res 22:5456–5465

Kostyu DD, Pfhol J, Ward FE, Lee J, Murray A, Amos DB (1993) Rapid HLA-DR oligotyping by an enzyme-linked immunosorbent assay performed in microtiter trays. Hum Immunol 38: 148–158

Saiki RK, Bugawan TL, Horn GT, Mullis KB, Erlich HA (1986) Analysis of enzymatically amplified beta-globin and HLA-DQ alpha DNA with allele-specific oligonucleotide probes. Nature 324:163–166

Southern EM, Maskos U, Elder JK (1992) Analyzing and comparing nucleic acid sequences by hybridization to arrays of oligonucleotides: evaluation using experimental models. Genomics 13:1008–1017

Southern EM, Case-Green SC, Elder JK, Johnson M, Mir KU, Wang L, Williams JC (1994) Arrays of complementary oligonucleotides for analysing the hybridisation behaviour of nucleic acids. Nucleic Acids Res 22:1368–1373

Syvanen AC, Sajantilla A, Lukka M (1993) Identification of individuals by analysis of biallelic DNA markers, using PCR and solid phase minisequencing. Am J Hum Genet 52:46–59

Weir BS, Triggs CM, Starling L, Stowell LI, Walsh KAJ, Buckleton J (1997) Interpreting DNA mixtures. J Forensic Sci 42: 213–222